EXHIBIT N

Exhibit 6 – Tong

'156 Patent

Claim Limitation (Claim 7)	Exemplary Disclosure
[**156a**] A device comprising:	Tong discloses a device. Specifically, Tong concerns the power consumption of floating point hardware in computer processors, which in turn are integrated into larger computing devices. *See, e.g.*: "FLOATING-POINT (FP) hardware provides a wide dynamic range of representable real numbers, freeing programmers from writing the cumbersome manual scaling code required of fixed-point representation." Tong at 273. "Section III constructs an ASIC-style FP multiplier for use as a baseline in subsequent experiments, and characterizes the power dissipation of its component pieces." Tong at 273. "[W]e have adopted a standard ASIC style design methodology and constructed a complete—though minimally optimized—FP multiplier." Tong at 274. "As a result of these experiments, we suggest that the fact that some operands need different bitwidths should be exposed to compilers. This would allow us to attempt compile-time precision/range optimizations which could then be realized using either variable-precision or a collection of mixed-precision arithmetic units in hardware." Tong at 284–85. "We argue that FP computation is inevitable as low-power systems evolve and that studies of the tradeoffs available among operand precision, dynamic range, and rounding modes can play an important role here." Tong at 285.
[**156b**] at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second	Tong discloses at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See, e.g.*: "This paper explores ways of reducing FP power consumption by minimizing the bitwidth representation of FP data. Analysis of several FP programs that manipulate low-resolution human sensory data shows that these programs suffer no loss of accuracy even with a

Exhibit 6 – Tong

Claim Limitation (Claim 7)	Exemplary Disclosure
numerical value,	significant reduction in bitwidth. Most FP programs in our benchmark suite maintain the same output even when the mantissa bitwidth is reduced by half. This FP bitwidth reduction can deliver a significant power saving through the use of a variable bitwidth FP unit. Our results show that up to 66% reduction in multiplier energy/operation can be achieved in the FP unit by this bitwidth reduction technique without sacrificing any program accuracy." Tong at 273 (Abstract). "FLOATING-POINT (FP) hardware provides a wide dynamic range of representable real numbers, freeing programmers from writing the cumbersome manual scaling code required of fixed-point representation." Tong at 273. "[F]or an increasing number of embedded applications, such as voice recognition and image/vision processing, FP hardware's performance, simplified programming model, and adaptability over a wide dynamic range makes it a desirable feature." Tong at 273. "One important characteristic of many of these mobile/portable applications is that they work primarily with low-resolution (4–10 bits) human sensory data, notably digitized speech and images. FP arithmetic allows many signal processing algorithms to process these acquired voice or vision data without concern about either the precision or range of intermediate results. Indeed, it remains common design practice [26] to prototype these algorithms in FP, then (usually manually) to translate them into an appropriate set of integer operations. In this study, we suggest that instead of rendering." Tong at 273. "Specifically, we examine how software can employ the ***minimal*** number of mantissa and exponent bits in FP hardware to ***reduce*** power consumption, yet ***maintain*** a program's overall accuracy. We study experimentally the relationship between the accuracy of FP programs and the number of bits used in the representation of their data. Our central results demonstrate that many programs which manipulate human sensory inputs, e.g., speech and image recognition, suffer no loss of accuracy with reduced bitwidth in the mantissa or exponent. We use this result to explore various methods to minimize power dissipation in the FP unit. The fundamental principle is to reduce waste—in this case, unnecessary bits in the FP representation and computation." Tong at 273 (emphasis in original).

Exhibit 6 – Tong

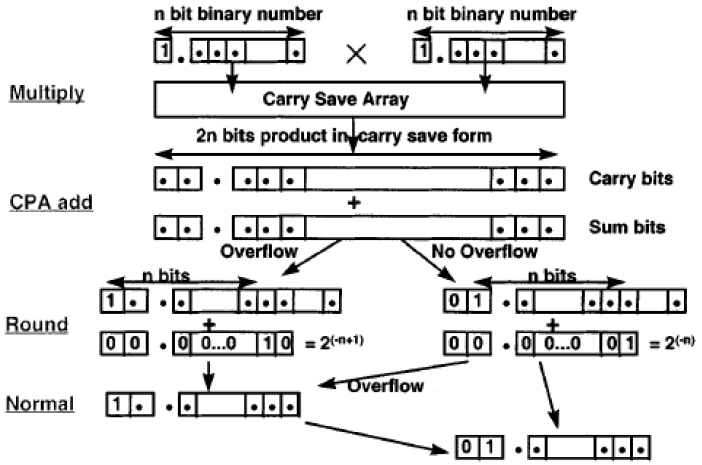
Claim Limitation (Claim 7)	Exemplary Disclosure
	"Section III constructs an ASIC-style FP multiplier for use as a baseline in subsequent experiments, and characterizes the power dissipation of its component pieces." Tong at 273. "For an increasing number of embedded applications such as voice recognition, vision/image processing, and other human-sensory-oriented signal-processing applications, FP's simplified programming model (in contrast with fixed-point systems) and large dynamic range makes FP hardware a desirable feature. Furthermore, many recognition algorithms achieve a high degree of accuracy starting from fairly low-resolution input sensory data. Leveraging these characteristics by allowing software to use the minimal number of mantissa and exponent bits, standard FP hardware can be modified to significantly reduce its power consumption while maintaining a program's overall accuracy." Tong at 274. "[W]e have adopted a standard ASIC style design methodology and constructed a complete—though minimally optimized—FP multiplier." Tong at 274.  Fig. 3. Mantissa data flow of an FP multiplier. Tong at Fig. 3.

Exhibit 6 – Tong

Claim Limitation (Claim 7)	Exemplary Disclosure
	"The most intuitive way of reducing the complexity of FP operations will be to reduce the number of bits in the FP representation. According to [1] and [9], a multiplier's power consumption decreases rapidly with the operand bitwidth." Tong at 277. "Consider again a typical FP format: one sign bit, a relatively large fractional mantissa field, a relatively smaller exponent field. Since the sign is only 1 bit, we only examine the possibility of reducing the number of bits used in the exponent and the mantissa. As the following sections reveal, reduction in the mantissa bitwidth is the most effective means of reducing power dissipation in FP datapath elements." Tong at 277. "It is obvious that power dissipation in an FP unit can be reduced by using fewer bits in the FP representation. However, fewer bits reduces precision and might result in a less accurate output. In this section, we first measure empirically the bitwidth requirements of several real programs, then suggest techniques for minimizing power dissipation by using smaller bitwidth functional units." Tong at 278. "These results clearly demonstrate that not all programs need the precision provided by generic FP hardware. The reason behind this result is that many programs dealing with human interfaces process sensory data with intrinsically low resolutions." Tong at 278.
[**156c**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented	Tong discloses the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. Specifically, Tong discloses experimental results teaching that for certain applications the optimum balance of precision and power consumption can be achieved using mantissa bitwidth reduction and exponent bitwidth that meet the claimed error rates and dynamic ranges, respectively. *See, e.g.*: "Our results show that programs such as speech recognition and image processing use

Exhibit 6 – Tong

Claim Limitation (Claim 7)	Exemplary Disclosure
by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input; and	significantly less power with our reduced bitwidth FP representation than with an IEEE-standard FP representation. For example, the speaker-independent continuous-speech recognition program Sphinx [10] requires only 5 mantissa bits and 6 exponent bits to maintain 90% recognition accuracy—the same level of accuracy achieved with a 32-bit IEEE-standard FP representation. Compared to a conventional 32-bit FP multiplier, which uses 2078 pJ per multiplication in our implementation, our reduced bitwidth FP multiplier uses only 705 pJ, roughly one third the energy. It has long been known that many such signal processing applications can get by with less precision/range than full FP. Our central contribution here is precise documentation of the extremity of such allowable reductions across several real benchmarks. In many cases, fewer the half the bits, in an appropriate custom FP format, suffice." Tong at 273. "To study the relationship between program accuracy and number of bits in FP representation, we have collected a set of five signal processing applications. These programs mainly deal with human sensory data such as digitized images and speech. All of them are single-precision FP programs and they are described in Table IV." Tong at 278. "To determine the impact of different mantissa and exponent bitwidths, we emulated in software different bitwidth FP units by replacing each FP operation with a corresponding function call to our FP software emulation package that initially implements the IEEE-754 standard (Fig. 5). Careful modifications to the FP emulation package allowed us to emulate different mantissa and exponent bitwidths. Then, each program was run using the modified FP package, and the results were compared to determine application accuracy."

Exhibit 6 – Tong

Claim Limitation (Claim 7)	Exemplary Disclosure

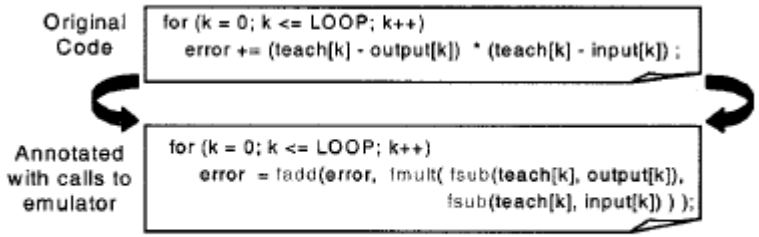


Fig. 5. FP emulation by annotating the source code.

Tong at 278 & Fig. 5.

"Fig. 6 plots the accuracy for each of the five programs across a range of mantissa bitwidths. None of the workloads display a noticeable degradation in accuracy when the mantissa bitwidth is reduced from 23 to 11 bits. For ALVINN and Sphinx III the results are even more promising; the accuracy does not change significantly with as few as 5 mantissa bits.

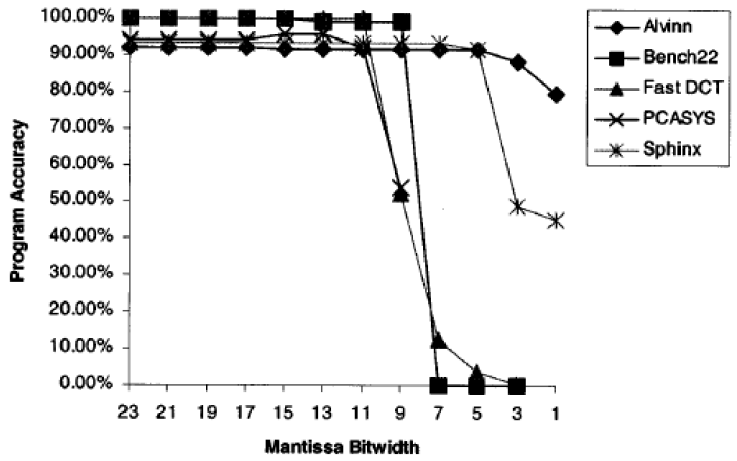


Fig. 6. Program accuracy across various mantissa bitwidths.

Tong at 279 & Fig. 6.

Exhibit 6 – Tong

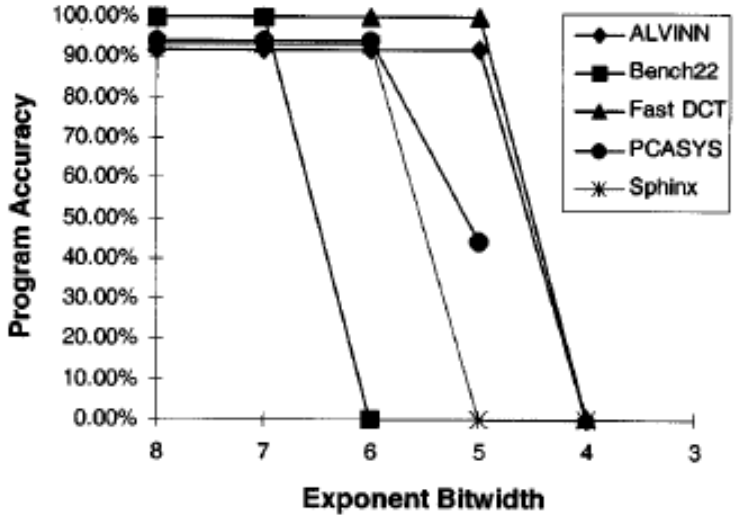
Claim Limitation (Claim 7)	Exemplary Disclosure
	"Fig. 7 shows that each program's accuracy has a similar trend when the exponent bitwidth is varied. With seven or more exponent bits, the error rates remain quite stable. Once the exponent bitwidth drops below six, the error rates increase dramatically and in some cases the programs could not finish properly." Tong at 278.  Fig. 7. Program accuracy across various exponent bitwidths. Figs. 6 and 7 show that we can reduce both the mantissa and exponent bitwidth without affecting the accuracy of the programs. This effect is especially prominent in the mantissa. This reduction of bitwidth can be turned into a reduction in power dissipation with the use of appropriate arithmetic circuits. Tong at Fig. 7. "What is quite clear from these experiments is that the FP format provides essential dynamic range (we can reduce, but not reduce dramatically, the number of exponent bits) but the fine precision of the 23-bit mantissa is not essential (half as many bits often suffice)." Tong at 279.

Exhibit 6 – Tong

Claim Limitation (Claim 7)	Exemplary Disclosure
	"But for the workloads in Table IV, three programs require 6 bits or more in the exponent to preserve a reasonable degree of accuracy, which means they need more than the typical 32 bits of precision that fixed point arithmetic offers. Simply using fixed-point representation without additional (usually manual) scaling will not resolve the problem." Tong at 279. "It should be noted that these complex applications were aggressively tuned by various software designers to achieve good performance using full IEEE representation. However, Figs. 6 and 7 show that significantly smaller bitwidth FP units can be used without compromising the necessary accuracy. For instance, certain FP constants in the Sphinx III code required more than 10 bits of mantissa to represent, but we modified those numbers so they could be represented using fewer bits during our experiment. Nevertheless (and perhaps somewhat surprisingly), these reductions in precision had little impact on the overall speech recognition accuracy. We believe that if the numerical behavior of these applications were tuned explicitly to a smaller bitwidth FP unit, we could obtain even better performance." Tong at 279. "For 8-bit multiplication, the digit-serial multiplier consumes 78% less energy than the 24 24Wallace tree multiplier (in the case of Sphinx and ALVINN). When 9–16 bits of the mantissa are required (in the case of Fast DCT, PCASYS and Bench22), the digit-serial multiplier still consumes 32% less energy than the 24 24 Wallace tree multiplier." Tong at 281 (disclosing hardware optimized for using a mantissa of 8 bits or less).

Exhibit 6 – Tong

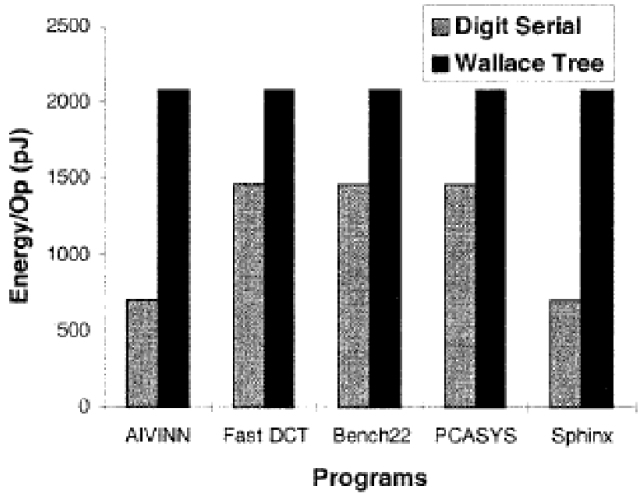
Claim Limitation (Claim 7)	Exemplary Disclosure
	 Fig. 11. Estimated multiplier energy reduction using digit-serial multiplier. A reduction of up to 66% in energy/op is attainable for Sphinx and ALVINN with the use of a digit-serial multiplier. Both Sphinx and ALVINN need only 5 bits of mantissa to be 90% accurate, and thus an 8-bit operand bitwidth is used for the digit-serial multiplier. PCASYS requires 11 bits of mantissa while Bench22 requires 9 bits, and thus a 16-bit operand bitwidth is used which results in a 30% energy/op reduction. Tong at 281 and Fig. 11. "Both Sphinx and ALVINN need only 5 bits of mantissa to be 90% accurate, and thus an 8-bit operand bitwidth is used for the digit-serial multiplier." Tong at 282. "[S]ince Sphinx requires only 5 bits of mantissa and 6 bits of exponents to maintain 90% of recognition accuracy, the programmer can simply choose an appropriate, available lower precision FP format." Tong at 282. *See also* Appendix to Responsive Contentions Regarding Non-Infringement and Invalidity ("Responsive Contentions") (detailing error rates associated with different mantissa sizes).

Exhibit 6 – Tong

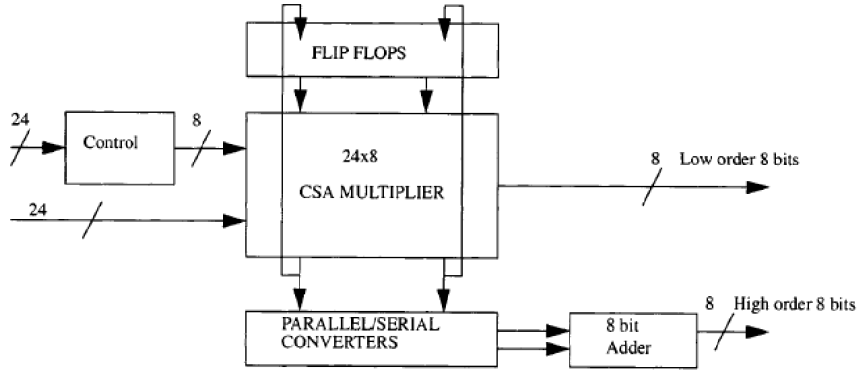
Claim Limitation (Claim 7)	Exemplary Disclosure
[**156d**] at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit;	Tong discloses at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit. As noted above, Tong discloses a low precision high dynamic range execution unit that a person of ordinary skill in the art would have understood to be integrated into a larger computer system. *See* [**156a**]–[**156b**]. Such a system included a computing device adapted to control the operation of an LPHDR execution unit. *See, e.g.*: To perform accurate comparisons, a complete 24 8 digit-serial multiplier was modeled in Verilog and then taken to layout using our previously described ASIC design flow, using a standard 0.5- m CMOS process. Synopsys' Design Compiler tool was used to synthesize the multiplier's control logic; the complete physical design was done using Duet's Epoch layout tool." Tong at 280. "The lower precision digit-serial design is slightly larger, since it includes control logic, and was also created in an ASIC style, in contrast to the larger Wallace tree, which is from an optimized layout compiler. Nevertheless, the digit-serial design is considerably faster, since the lowered precision (8 24 versus 24 24) creates shallower logic." Tong at 280.  Fig. 9. Block diagram of a 24 × 8 digit-serial multiplier. Tong at 281 and Fig. 9. To the extent that Singular contends that Tong does not disclose a controller, notwithstanding

Exhibit 6 – Tong

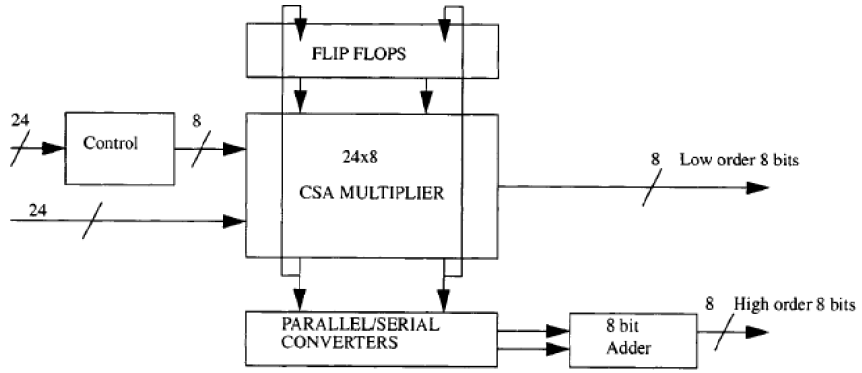
Claim Limitation (Claim 7)	Exemplary Disclosure
	these disclosures, a person of ordinary skill in the art would have understood the larger computer system to contain a computing device adapted to control the operation of the low precision high dynamic range execution unit. *See* Responsive Contentions.
[**156e**] wherein the at least one first computing device comprises at least one of a central processing unit (CPU), a graphics processing unit (GPU), a field programmable gate array (FPGA), a microcode-based processor, a hardware sequencer, and a state machine;	Tong discloses at least one first computing device comprises at least one of a central processing unit (CPU), a graphics processing unit (GPU), a field programmable gate array (FPGA), a microcode-based processor, a hardware sequencer, and a state machine. Specifically, Tong discloses that the controller for its floating-point multiplier is a finite state machine: "[W]e explore in this section one architecture for a variable bitwidth FP multiplier to reduce power consumption. To support variable bitwidth multiplications (up to 24 x 24 bit), we considered a 24 x 8 bit *digit-serial* architecture similar to the one described in Hartley and Parhi [12]. The 24 x 8 bit architecture (see Fig. 9) allows us to perform 8-, 16-, and 24-bit multiplication by passing the data once, twice, or three times though [*sic*] the serial multiplier. A finite state machine is used to control the number of iterations through the core multiplier." Tong at 280 (emphasis in original).  Fig. 9. Block diagram of a 24 × 8 digit-serial multiplier. Tong at Fig. 9. To the extent that Singular contends that Tong does not disclose this limitation, notwithstanding

Exhibit 6 – Tong

Claim Limitation (Claim 7)	Exemplary Disclosure
	its explicit disclosure of a finite state machine to control the number of iterations through the core multiplier, a person of ordinary skill in the art would have understood that Tong discloses a low precision high dynamic range execution unit that would be integrated into a larger computer system. *See* [**156a**]–[**156b**]. As also noted above, a person of ordinary skill in the art would have understood the larger computer system to contain a computing device adapted to control the operation of the low precision high dynamic range execution unit. *See* [**156d**]. A person of ordinary skill in the art would have understood the computing device could, among other things, be a CPU. *See* Responsive Contentions.
[**156f**] and, wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide.	Tong discloses the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide. Specifically, Tong recognizes that the benefits of power optimization through mantissa bitwidth reduction could be realized "using either variable-precision or a collection of mixed-precision arithmetic units in hardware." Tong at 285. To the extent Singular contends that Tong does not disclose this limitation, the limitation would have been obvious to a person of ordinary skill in the art as any person of ordinary skill in the art would have been motivated to achieve the power-optimization described by Tong in a system with large numbers of low-precision high-dynamic range units, such as massively parallel processing supercomputers with large numbers of processing elements configured to perform floating-point arithmetic or personal computers utilizing a single-instruction, multiple datastream architecture. *See* Responsive Contentions.

Exhibit 6 – Tong

'273 Patent

Claim Limitation (Claim 53)	Exemplary Disclosure
[**273a**] A device:	Tong discloses a device. *See* [**156a**].
[**273b**] comprising at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	Tong discloses at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**].
[**273c**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input;	Tong discloses the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. *See* [**156c**].

Exhibit 6 – Tong

Claim Limitation (Claim 53)	Exemplary Disclosure
[**273d**] wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide.	Tong discloses the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide. *See* [**156f**]. To the extent Singular contends that Tong does not disclose this limitation, the limitation would have been obvious to a person of ordinary skill in the art as any person of ordinary skill in the art would have been motivated to achieve the power-optimization described by Tong in a system with large numbers of low-precision high-dynamic range units, such as massively parallel processing supercomputers with large numbers of processing elements configured to perform floating-point arithmetic or personal computers utilizing a single-instruction, multiple datastream architecture. *See* Responsive Contentions

Exhibit 6 – Tong

'961 Patent

Claim Limitation (Claim 4)	Exemplary Disclosure
[**961a**] A device comprising:	Tong discloses a device. *See* [**156a**].
[**961b**] at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	Tong discloses at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**].
[**961c**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input; and	Tong discloses the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. *See* [**156c**].
[**961d**] at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit.	Tong discloses at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit. *See* [**156d**].

Exhibit 6 – Tong

Claim Limitation (Claim 13)	Exemplary Disclosure
[**961e**] A device comprising:	Tong discloses a device. *See* [**156a**].
[**961f**] a plurality of components comprising:	Tong discloses a plurality of components. *See* [**156b**] + **[156d**].
[**961g**] at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	Tong discloses at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**].
[**961h**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input.	Tong discloses the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. *See* [**156c**].